

## **APPENDIX**

### **Explanatory Material**

#### **SAMPLE DESIGN AND ESTIMATION PROCEDURES**

The current sample was introduced with the release of QSS estimates for the fourth quarter of 2006. This sample was designed to produce estimates based on the 2002 North American Industry Classification System (NAICS). This section describes the design, selection, and estimation procedures for the new sample. For descriptions of prior samples, see the *Quarterly Services Survey* publications.

#### **Sampling Frame**

The sampling frame for the Quarterly Services Survey sample is a subset of the Service Annual Survey sample and has the same types of sampling units as the Service Annual Survey frame—large, multiple establishment firms and Employer Identification Numbers (EINs). The EIN is the identifier that employer businesses use to report Social Security payroll withholdings to the federal government. Both sampling units represent clusters of one or more establishments owned or controlled by the same firm.

#### **Stratification, Sampling Rates, and Allocation**

The primary stratification of the Quarterly Services Survey frame is by industry group based on the detail required for publication. We publish Quarterly Services Survey estimates at broader industry levels than the industry groupings used to publish Service Annual Survey estimates. Therefore, the industry stratification for the Quarterly Services Survey sample is broader than the industry stratification used for the Service Annual Survey sample.

Within industry group, we substratify the sampling units by a measure of size related to their annual revenue as reported in the Service Annual Survey. We select sampling units expected to have a large effect on the precision of the estimates “with certainty.” This means they are sure to be selected and will represent only themselves (i.e., have a selection probability of 1 and a sampling weight of 1). To identify the certainty units, we determine a substratum boundary (or cutoff) that divides the certainty units from the noncertainty units. We base these cutoffs on a statistical analysis of data from the 2002 Economic Census. Accordingly, these values are on a 2002 revenue basis. We also use this analysis to determine the number of size substrata for each industry group and to set preliminary sampling rates needed to achieve specified sampling variability objectives on revenue estimates for different industry groups. The size substrata and sampling rates are later updated through analysis of the data extracted from the Census Bureau’s Business Register.

#### **Sample Selection**

We select the Quarterly Services Survey sample independently within each size substratum contained in an industry stratum. We select firms with certainty if the estimated annual revenue exceeds the certainty cutoff. The selection procedure for the noncertainty portion of the sample follows a systematic, probability proportional-to-size scheme. Because the Quarterly Services Survey sample is an independently selected subsample, it is possible that we select some units in the Service Annual Survey sample at a lower sampling rate than desired for the Quarterly Services Survey sample. We include such a unit in the Quarterly Services Survey sample and assign a sampling weight equal to the unit's Service Annual Survey sampling weight. The maximum sampling weight for an EIN selected for the Quarterly Services Survey sample is about 750.

### **Sample Maintenance**

Periodically, we update the Quarterly Services Survey sample to represent EINs issued since the initial sample selection. These new EINs, called births, are EINs recently assigned by the Internal Revenue Service (IRS) on the latest available IRS mailing list for Federal Insurance Contributions Act taxpayers and assigned an industry classification (if possible) by the Social Security Administration.

We sample EIN births on a quarterly basis using a two-phase selection procedure. To be eligible for selection, a birth must either have no industry classification or be classified in an industry within the scope of one of the following: the Annual Retail Trade Survey, the Annual Trade Survey, or Service Annual Survey, and it must meet certain criteria regarding its number of paid employees or quarterly payroll. In the first phase, we stratify births by industry and a measure of size based on quarterly payroll. A relatively large sample is drawn and canvassed to obtain a more reliable measure of size, consisting of sales in 2 recent months, company affiliation information, and a new or more detailed industry classification code.

Using this more reliable information, we subject the selected births from the first phase to probability proportional-to-size sampling with overall probabilities equivalent to those used in drawing the initial Annual Retail Trade Survey, Annual Trade Survey, and Service Annual Survey samples from the December 2004 Business Register. Because of the time it takes for a new employer firm to acquire an EIN from the IRS, and because of the time needed to accomplish the two-phase birth-selection procedure, we add births to the samples approximately 9 months after they begin operation.

Updates to the Quarterly Services Survey sample occur in the same manner and at the same time as updates to the Service Annual Survey sample. The births selected for the Quarterly Services Survey sample are a subset of the births selected for the Service Annual Survey sample. These births are selected using sampling rates equivalent to those used in selecting the initial Quarterly Services Survey sample.

### **Estimation Method**

Direct expansion estimates for each quarter are computed as the sum of weighted data (reported or imputed), where the weight for each unit is the reciprocal of the probability of selection of the unit into the sample. The direct expansion estimates are input into the benchmarking procedure to produce the published estimates, as described in the **Revisions to Previously Published Quarterly Estimates** section of this report.

## **RELIABILITY OF THE ESTIMATES**

The published estimates may differ from the actual, but unknown, population values. For a particular estimate, statisticians define this difference as the total error of the estimate. When describing the accuracy of survey results, it is convenient to discuss total error as the sum of sampling error and nonsampling error. Sampling error is the error arising from the use of a sample, rather than a census, to estimate population values. Nonsampling error encompasses all other factors that contribute to the total error of a survey estimate. The sampling error of an estimate can usually be estimated from the sample; whereas, the nonsampling error of an estimate is difficult to measure and can rarely be estimated. Consequently, the actual error in an estimate exceeds the error that can be estimated. Further descriptions of sampling error and nonsampling error are provided in the following sections. Data users should take into account the estimates of sampling error and the potential effects of nonsampling error when using the published estimates.

### **Sampling Error**

Because the estimates are based on a sample, exact agreement with results that would be obtained from a complete enumeration of the sampling frame using the same enumeration procedures is not expected. However, because each firm on the sampling frame has a known probability of being selected into the sample, it is possible to estimate the sampling variability of the survey estimates.

The particular sample used in this survey is one of a large number of samples of the same size that could have been selected using the same design. If all possible samples had been surveyed under the same conditions, an estimate of a population parameter of interest could have been obtained from each sample. For the parameter of interest, estimates derived from the different samples would, in general, differ from each other. Common measures of the variability among these estimates are the sampling variance, the standard error, and the coefficient of variation (CV). The sampling variance is defined as the squared difference, averaged over all possible samples of the same size and design, between the estimator and its average value. The standard error is the square root of the sampling variance. The CV expresses the standard error as a percentage of the estimate to which it refers. For example, an estimate of 200 units that has an estimated standard error of 10 units has an estimated CV of 5 percent. The sampling variance, standard error, and CV of an estimate can be estimated from the selected sample because the sample was selected using probability sampling. Note that measures of sampling variability, such as the standard error and CV, are estimated from the sample and are also subject to sampling variability. (Technically, we should refer to the *estimated* standard error or the *estimated*

CV of an estimator. However, for the sake of brevity we have omitted this detail.) It is important to note that the standard error and CV only measure sampling variability. They do not measure any systematic biases in the estimates.

The Census Bureau recommends that individuals using published estimates incorporate this information into their analyses, as sampling error could affect the conclusions drawn from these estimates.

The estimate from a particular sample and its associated standard error can be used to construct a confidence interval. A *confidence interval* is a range about a given estimator that has a specified probability of containing the average of the estimates derived from all possible samples. Associated with each interval is a percentage of confidence, which is interpreted as follows. If, for each possible sample, an estimate of a population parameter and its approximate standard error were obtained, then:

1. For approximately 90 percent of the possible samples, the interval from 1.65 standard errors below to 1.65 standard errors above the estimate would include the average of the estimates derived from all possible samples of the same size and design.
2. For approximately 95 percent of the possible samples, the interval from 1.96 standard errors below to 1.96 standard errors above the estimate would include the average of the estimates derived from all possible samples of the same size and design.

To illustrate the computation of a confidence interval for an estimate of total revenue, assume that an estimate of total revenue is \$10,750 million and the CV for this estimate is 1.8 percent, or 0.018. First obtain the standard error of the estimate by multiplying the total revenue estimate by its CV. For this example, multiply \$10,750 million by 0.018. This yields a standard error of \$193.5 million. The upper and lower bounds of the 90-percent confidence interval are computed as \$10,750 million plus or minus 1.65 times \$193.5 million. Consequently, the 90-percent confidence interval is \$10,431 million to \$11,069 million. If corresponding confidence intervals were constructed for all possible samples of the same size and design, approximately 9 out of 10 (90 percent) of these intervals would contain the average of the estimates derived from all possible samples.

## **Nonsampling Error**

Nonsampling error encompasses all other factors, other than sampling error, that contribute to the total error of a sample survey estimate and may also occur in censuses. It is often helpful to think of nonsampling error as arising from deficiencies or mistakes at some point in the survey process. Nonsampling errors are difficult to measure and can be attributed to many sources: the inclusion of erroneous units in the survey (overcoverage), the exclusion of eligible units from the survey (undercoverage), nonresponse, misreporting, mistakes in recording and coding responses, misinterpretation of questions, and other errors of collection, response, coverage, or processing. Additional

nonsampling error may have been introduced by the method used to adjust the survey estimates using results of the 2002 Economic Census. Although nonsampling error is not measured directly, the Census Bureau employs quality control procedures throughout the process to minimize this type of error.

A potential source of bias in the estimates is nonresponse. Nonresponse is defined as the inability to obtain all the intended measurements or responses about all selected units. Two types of nonresponse are often distinguished. *Unit nonresponse* is used to describe the inability to obtain any of the substantive measurements about a sampled unit. In most cases of unit nonresponse, the questionnaire was never returned to the Census Bureau after several attempts to elicit a response. *Item nonresponse* occurs either when a question is unanswered or the response to the question fails computer or analyst edits.

For both unit and item nonresponse, a missing value is replaced by a predicted value obtained from an appropriate model for nonresponse. This procedure is called *imputation* and uses survey data and administrative data as input. For NAICS 51 (Information sector), imputed revenue amounts to about 10 percent of the total revenue estimate; for NAICS 54 (Professional, scientific, and technical services) imputed revenue amounts to about 25 percent; for NAICS 56 (Administrative and support and waste management and remediation services) imputed revenue amounts to about 30 percent; and for NAICS 62 (Hospitals, and nursing and residential care facilities) imputed revenue amounts to about 20 percent of total revenue.

## DEFINITION OF TERMS

**Establishment.** A single physical location where business is conducted or services are performed.

**Firm.** A business organization or entity consisting of one or more domestic establishments/locations under common ownership or control.

**Total expenses.** (Basic dollar volume measure of expenses for firms exempt from federal income tax.) Costs incurred during the survey year whether or not payments were made in that year. Total expenses include annual payroll; employee benefits, interest, and rent expenses; supplies used for operating; cost of merchandise sold; and other expenses allocated to operations during the year. Also included are contracted or purchased services; fees paid to other organizations for fundraising; depreciation expenses; and expenses of locations providing support services (e.g., repair services, administrative services, etc.) for service establishments. Total expenses exclude outlays for the purchase of real estate (land and buildings); for construction; for additions, major alterations, and improvements to existing facilities; all other capital expenditures; funds invested; income taxes; assessments (dues) paid to the parent or other chapters of the same organization. Firms engaged in raising funds, exclude funds that are transferred to charities or other organizations.

**Federal income tax status.** Firms that indicate all or part of their income are exempt from federal income tax under provision of Sections 501 or 521 of the Internal Revenue Service are classified as tax-exempt. Firms indicating no such exemption are classified as taxable. For all firms, the tax status classification is based upon administrative records.

**Total operating expenses.** Costs incurred during the survey year, even though payment may be made at a later date. Excludes interest on loans and sales taxes and other taxes collected from customers and paid directly to a taxing authority.

**Total operating revenue.** Includes charges or billings for services rendered and any sales of merchandise during the survey year, even though payments may be received at a later date. Excludes income from interest, investments, gifts, loans, contributions, or grants; the sale of securities, real estate, etc; sales taxes or other taxes collected from customers and remitted directly by the firm to a local, state, or federal tax agency; revenue from the sale of merchandise and equipment from retail establishments; and revenue from a domestic parent organization, or from franchise locations owned by others and any franchise or license fees.

**Total revenue.** (Basic dollar volume measure for firms exempt from federal income tax.) Charges or billings to customers or clients for services rendered and merchandise sold during the survey year whether or not payment was received in that year. Also includes income from interest, dividends, contributions, gifts and grants, rents, royalties, dues and assessments from members and affiliates, and net receipts from fundraising activities. Receipts from taxable business activities, as well as tax-exempt activities are included. Excludes sales taxes or other taxes (real estate, admissions, etc.) collected by the organization from customers or clients and paid directly to local, state, or federal income tax agencies; income from the sale of real estate, investments, or other assets; or amounts transferred to operating funds from capital or reserve funds. Firms providing legal services report payments received in the survey year regardless of when services are rendered.